Satellite EO Design Report - Draft

Satellite Monitoring for Forest Management (SMFM) Project

27 April 2018

LTS International Limited (UK)
Pentlands Science Park
Bush Loan
Penicuik
Nr. Edinburgh EH26 0PL
Scotland

(+44) 131 440 5500

(+44) 131 440 5501

mail@ltsi.co.uk

@LTS_Int

www.ltsi.co.uk

Registered in Scotland Number 100833

Table of Contents

Executive Summary

1. Tropical dry forests are subject to some of the highest rates of deforestation and degradation around the world. These ecosystems are particularly at risk due to their fragility and the high demand for forest goods and services, which are required to support the livelihoods of large numbers of the world's poorest people. The need for dry forest management has also required enhanced methods and approaches for monitoring forest change and assessment of deforestation and degradation. Satellite Earth Observation (EO) provides unprecedented views of the Earth and the data generated present an opportunity to address the existing limitations in forest monitoring capabilities at large scales.

2. Most existing EO-based forest monitoring methods have been developed for moist tropical forest ecosystems, though these do not always work well in the less studied and more challenging tropical dry forest ecosystems. There are opportunities for improved monitoring techniques for dry forests in increasing volumes of open access data, cloud processing capabilities and the new satellite system's performance capabilities (e.g. wider spectral range, long-wavelength radar, higher spatial resolution, increased frequency of data acquisition and pre-processed data products) for dry forest monitoring and management. By leveraging this new data, the SMFM project aims to improve methods for EO monitoring of dry forests.

3. The Satellite EO Design report outlines the functionality and data requirements of the EO tools developed for the project. The tools are designed to support land use and forest monitoring and management. The report outlines the methodologies and technologies used to produce, including new or enhanced satellite EO methods to address the requirements and gaps in capabilities defined in Task 1. The plan for the application of new satellite EO data (e.g. Sentinel-2, Sentinel-1) for monitoring tropical dry forests and forest degradation assessment based on related products frameworks (e.g. GFOI Method and Guidance Document) is also described and the satellite data acquisition plan further considered. The report includes the next steps in tool development in relation to the work programme and contribution towards the project objectives. The information and analysis in this report is particularly relevant to the design and implementation of associated capacity building activities.

4. The tools will include an approach to LU/LC mapping using a fusion of Sentinel-1 (S1) radar and Sentinel-2 (S2) optical data. This will explore the use of multi-temporal S1 and S2 images to produce LU/LC classification maps in dry forest ecosystems, and will provide a semi-automated processing chain for data pre-processing and analysis based on a cloud-based platform. Using dense time-series of S1 and S2 data, the tools will aim to detect deforestation events and provide a near-real-time early warnings of deforestation events, and where feasible estimates of degradation.

5. For forest biomass and degradation mapping, the SMFM project will produce a radar L-band based tool, which will be able to provide the crucial historical data on biomass stocks, degradation, and deforestation from the Advanced Land Observing Satellite (ALOS-1/ALOS-2). The methods used will be compatible with upcoming L-band missions, which will provide free data (e.g. NASA-ISRO Synthetic Aperture Radar (NISAR) and Satellites for Observation and Communications (SAOCOM)). The outputs and data produced from an L-band approach will then be used as a forest biomass and degradation estimation baseline, and be compared to the equivalent results using S1/S2 data.

6. Finally, a tool will be developed to identify the causes of forest change, based on properties of change events such as their size, shape and intensity. This tool will be calibrated with field data, and tested on outputs from S1/S2 and ALOS change detection methods.

1 Background

1.1 Introduction

7. In the SMFM Inception Report, the LTS Consortium and the World Bank agreed to the development of four Earth Observation (EO) methods ("tools") to assist in the satellite monitoring of dry forests. The proposed tools were based on the immediate and future needs of selected partner countries, the requirements of the World Bank, and the available remote sensing technologies for the monitoring of dry forest.

8. This report provides an overview of the EO methodologies and tools that are being developed under the SMFM project and contains a general review of the current EO methods for tropical dry forest monitoring, followed by an overview of the existing and upcoming cloud-based platforms for efficient data processing and storage, which assist in the processing and analysis of the large volume of satellite data now available.

9. The second part of the document describes the methodological approach of the four EO tools under development as part of the SMFM project. The first tool is presented in two sections, one for Sentinel-2 data and the other one for Sentinel-1 data. The tools presented in this report are:

- **Tool 1a:** Semi-automated pre-processing of Sentinel-2 data for LU/LC Classification
- **Tool 1b:** Semi-automated pre-processing of Sentinel-1 data for LU/LC Classification
- **Tool 2:** Annual forest biomass change and degradation mapping using the ALOS PALSAR mosaic
- **Tool 3:** Continuous forest change monitoring with Sentinel-2 data
- **Tool 4:** Identifying causes of forest change.

In references to tools within the report, the table below (Table 1), indicates what these are referred to as well as their purpose.

Table 1. SMFM tool references and description

Tool No.	Tool Name	Purpose
1a	sen2mosaic	Download, pre-processing and generation of cloud-free composite images and seamless mosaics for Setinel-2 images
1b	sen1mosaic	Download, pre-processing and generation of cloud-free composite images and seamless mosaics for Setinel-1 images
2	biota	Downloading, pre-processing of ALOS data, mapping aboveground biomass and forest cover, and detection of forest changes
3	deforest	Dense time-series of Sentinel-1 and 2 data for near-real time forest change monitoring
4	n/a	Classification of forest change type and identification of drivers of deforestation events

10. The analysis of in-country capacity and discussions with all stakeholders suggests that the best approach for developing the tools is to use free and open-sourced software. It appears that the technical capability to utilise the Linux command line and Python-based tools can be developed in both countries.

11. The tools will be formed of Python scripts and command-line tools, which can be implemented locally for smaller data processing tasks or on a cloud-based platform where required. These will be appropriately documented, both in the code and in supporting documents, training exercises, examples, and ultimately through the development of working practices and good practice guidelines.

12. For each tool the (i) objective, (ii) method design, (iii) user interface and usage, and (iv) documentation and distribution approach are presented. Whilst the methods for tools 1 and 2 are relatively far advanced, Tools 3 and 4 will involve further active research, so for these a research plan is presented.

1.2 Review of EO Methods for Dry Forest Monitoring

13. There are three main types of EO data: optical, radar and Lidar. Each has different characteristics, and different strengths and weaknesses for use in forest monitoring. For more and detailed information, refer to the SMFM Inception Report, Section 3.3.

1.2.1 Optical Remote Sensing

14. Optical remote sensing data is the most widely available form of EO data, with over 300[1] (and many more small to medium sized) satellites collecting data regularly. Optical data views the top of the forest canopy, and thus it can be used to assess canopy cover and potentially estimate the density and health of leaves. Optical data has been traditionally used to map deforestation and has been used for detection of forest degradation in wet tropical forests through a number of government programmes.

15. Whilst optical remote sensing data is widely used and methods are advanced, optical data has three major shortcomings when it comes to dry forest mapping:

- It cannot penetrate clouds, which cover much of the tropics most of the time, limiting observations.
- It cannot penetrate the top of the forest canopy, meaning low-level sub-canopy degradation, not involving canopy trees, will be invisible – this is especially a problem in dense woodlands and forests.
- Canopy leaves and grasses in the mixed tree-grass ecosystems that dominate the dry tropics are confused by optical sensors. Canopy and grass greenness varies dramatically over a year and between years, and limit the use of optical data to assess changes to just the woody component.

16. A range of methods exist for use of optical data for forest change mapping. The most straightforward is to compare single timestamp classified layers, though this can lead to somewhat greater inaccuracies due to the combination of errors in each layer. Time series analysis is used to assess temporal change of a parameter related to forest cover (e.g. normalised difference vegetation index) by analysing its trend through time and by identifying break-points (as the result, for example, of deforestation / degradation) or long-term trend (e.g. forest regrowth). Seasonal effects, differences in sensor calibration, atmospheric conditions and random error are all affecting each pixel's time series, therefore it is important that appropriate algorithms are developed to take those into account. Machine learning approaches have also been used to enhance the data analysis process, by using multiple data sets and/or metrics, and by using algorithms that are continually updated based on the best-available analytical approaches.

17. There are a number of options for free optical satellite data. The most widely used source of optical data for forest monitoring is Landsat, a series of satellites active since 1972 which acquire images every 16 days at 30 metres resolution. The NASA MODIS satellites offer greater temporal frequency of 1-2 days but at coarser resolution of 250 metres. The recently launched Sentinel-2 satellites, as part of the Copernicus Programme of ESA that guarantees data continuity through until 2030, offer unprecedented resolution (10 metres) and revisit times (6 days) for freely available data.

1.2.2 Radar Remote Sensing

18. Synthetic Aperture Radar (SAR) satellites are active sensors that look obliquely at the Earth's surface using microwave data (in the mm – cm wavelength range). Long-wavelength microwave data penetrates through the forest canopy to obtain information on forest structure, with the amount of radiation reflected back to the instrument ('backscatter') increasing as the number and/or size of trees present in an area increases. Radar satellites have been used to map aboveground biomass (AGB), and to map deforestation and degradation through AGB loss. Measurement of AGB with radar backscatter is limited by a saturation point, typically around 150 tonnes per hectare AGB for L-band (the longest, and therefore most sensitive, wavelength currently available), and ~50 tonnes per hectare for C-band (which is more widely available).

19. L-band Synthetic Aperture Radar has a long track record of use in mapping biomass and biomass change (degradation, regrowth) in dry forest and savanna regions of the tropics. There is a well-established relationship between L-band radar backscatter and AGB in moderate density woodlands and forest (e.g. Ryan et al. 2012), as shown in Figure 1. Up to the saturation point of ~150 tonnes per hectare in denser forests, this relationship is reasonably consistent between forest types (Mitchard et al. 2011), a property which has been

[1] https://www.ucsusa.org/nuclear-weapons/space-weapons/satellite-database#.WszUpS7waUI

used to map AGB in dry forest ecosystems over continental scales (Bouvet et al. 2018 & McNicol et al. *in press*). Dry forests can have biomass values up to 200 tonnes per hectare, but in the key countries for this project (Mozambique and Zambia) only a tiny proportion of forest is thought to be above the 150 tonnes per hectare saturation point for L-band radar (McNicol et al. *in press*), so this tool will only be applicable to the sparser woodlands and forests that are dominant across much of Mozambique and Zambia. Overall, there is a consensus in the scientific community that L-band cross-polarised radar represents the only space-borne technology proven to map forest AGB and AGB change in dry forest ecosystems (GFOI, 2014).

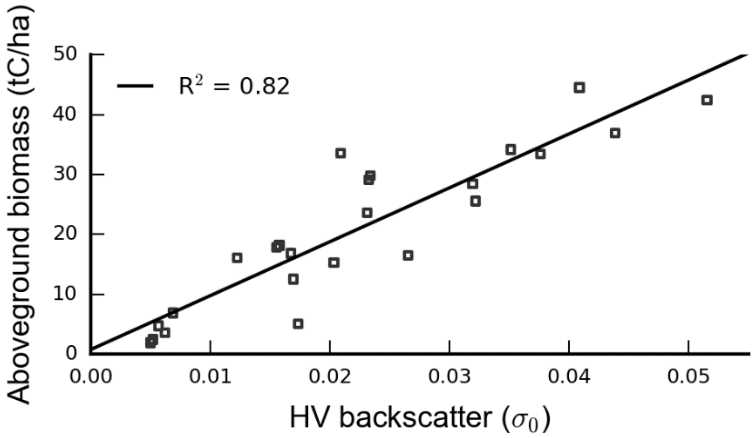

Figure 1: Biomass-backscatter plot showing the relationship between HV backscatter from ALOS PALSAR and a series of 1 hectare (100 x 100 metre) permanent forest plots from Miombo woodlands in Tanzania.

20. C-band radar data is more sensitive to differences in structure between vegetation types than L-band, more sensitive to grass biomass, and more sensitive to vegetation and soil moisture, therefore research has concentrated on L-band where data is available. There has however been success in using C-band to map forest type, for example the difference between natural forest, degraded forest and plantations (De Grandi et al 2015 & Dong et al 2015). Uptake of radar data in applications such as land cover mapping has been limited by a relatively complex pre-processing chain the limited availability of free data until recently.

21. The Japanese ALOS PALSAR-2, launched in 2014, is the only L-band radar satellite currently operational, but two additional L-band radar satellite missions SAOCOM[2] and NISAR[3] are planned to be launched in 2018 and 2020 respectively. An annual mosaic is produced by the Japanese Aerospace Exploration Agency (JAXA) using ALOS PALSAR/PALSAR-2 data at 25 metres resolution for the years 2007-2010 & 2015 onwards[4]. The ESA Sentinel-1 satellites from the Copernicus Programme provide, free of charge, C-band radar data at 10 metres resolutions from 2014 onwards.

1.2.3 Lidar Remote Sensing

22. Lidar EO satellites use laser light looking directly down to generate a profile of tree height and forest vertical structure. This enables a representation of a forest to be built up with structural parameters such as height, stem density and canopy cover to be estimated directly, and this data has been paired with optical data from MODIS to generate the first pan-tropical maps of aboveground biomass (Saatchi et al. 2011 & Baccini et al 2012). There are at present no operational satellite Lidar instruments.

23. Much more detailed data can be collected with Lidar instruments using aircraft of UAVs, with data detailed enough to identify individual trees. Repeat surveys can therefore see the removal of individual trees, and thus it is the only remote sensing method that can guarantee to map degradation with high accuracy even if the magnitude or size of disturbance is low. Airborne or UAV Lidar data are therefore the gold-standard for measuring forest degradation.

24. The high cost of collecting Lidar data from aircraft or UAVs and the absence of operational satellite instruments preclude the use of Lidar data for the mapping products of the SMFM project. There are however

[2] http://space.skyrocket.de/doc_sdat/saocom-1.htm;
[3] https://www.jpl.nasa.gov/missions/nasa-isro-synthetic-aperture-radar-nisar/
[4] http://www.eorc.jaxa.jp/ALOS/en/palsar_fnf/fnf_index.htm

two upcoming Lidar missions, ICESAT-2[5] and GEDI[6], which are planned to be launched in 2018 and 2019, respectively.

1.3 Cloud Platforms

25. Recent moves towards open access satellite data (e.g. Landsat, Sentinel-1/Sentinel-2) have hugely increased availability of data for forest monitoring, and provided opportunities to utilise dense time-series of satellite data to develop new methods to address the challenges associated with satellite monitoring of dry forest. Working with dense time series of satellite data at regional to national scales comes with computational challenges, particularly in locations where internet access is poor. Therefore, the SMFM project will make use of cloud platforms, where storage of large satellite archives is provided together with access to significant processing power on a remote server.

26. An overview of the available cloud-based platforms for the storage and processing of satellite imagery data on the cloud is given below. Once the SMFM tools are fully operational and tested, a thorough assessment will be made to identify the most suited platform for the implementation of these tools on the cloud. This will be based on criteria of country preferences, appropriateness, user-friendliness, degree of customization, and platform storage as well as processing costs. ESA's DIAS platform appears to have particular potential in regard to reaching a global audience, but other platforms will also be considered based on country partner needs. More details on the planning around this will be provided around June 2018.

1.3.1 SEPAL[7]

27. The System for Earth Observation Data Access, Processing and Analysis for Land Monitoring (SEPAL) is a collaboration between FAO and Norway. SEPAL is a cloud-based computing platform for geospatial big data processing and storage. SEPAL offers direct access to satellite data sources through a visual user interface, a set of open-source software tools, based on the Linux command line, and the capacity to run customized Python scripts and data processing chains on a virtual machine hosted by Amazon Web Services. SEPAL is not provided for free, but charges for both processing and storage.

1.3.2 EO IPT Poland[8]

28. The Earth Observation Innovative Platform Testbed Poland (EO ITP Poland) is a system from ESA being developed by the company CloudFerro. The service offers high performance computing with local access to large archive of the main EO datasets, including the Landsat and Sentinels satellites. Services are provided at a cost related to processing power as well as storage required.

1.3.3 CEOS Data Cubes[9]

29. The Committee on Earth Observation Satellites (CEOS) Data Cube Platform is a data processing platform for Earth science data, with a focus on remote-sensing data. The platform provides a data ingestion framework that includes support for automated ingestion of a wide variety of remote sensing data products. The data products are ingested into an N-dimensional data array. The Data Cube model aims to surpass the old data system model by providing standardised analysis ready data to users in a streamlined way, rather than unprocessed and uncorrected data that requires high network bandwidth, local processing infrastructure and local expertise. However, the goal is to have an operational CEOS platform by 2020, which is beyond the timescales of the SMFM project.

1.3.4 F-TEP[10]

30. Forestry Thematic Exploitation Platform (F-TEP) is an EO data processing and analysis platform under development by ESA. The aim is for a 'one-stop shop' of forestry remote sensing services for academic and commercial sectors. The service will offer large pre-processed satellite data archives, in addition to computing power and easy-to-use data processing tools and GIS software. The aim is to encourage the use of data from the Sentinel satellites support forest ecosystem monitoring and sustainable forest management. The project

---

[5] https://icesat-2.gsfc.nasa.gov/
[6] https://eospso.nasa.gov/missions/global-ecosystem-dynamics-investigation-Lidar
[7] https://sepal.io/
[8] https://eocloud.cloudferro.com/knowledgebase.php?action=displayarticle&id=13
[9] https://www.opendatacube.org/ceos
[10] https://forestry-tep.eo.esa.int/

is in a pilot phase, focussing on forest management in Mexico and Finland, and it is not currently open to general use.

1.3.5 DIAS[11]

31. ESA's Copernicus Data and Information Access Services (DIAS) platform will contain all the Copernicus EO satellite data and information and will offer free and open data download services to its users. However, infrastructure and operational costs are envisaged for could processing, and will be published before June 2018. DIAS integrates the principles of transparency of the services offer and the data offer, as well as an infrastructure for the development and execution of customisable scripts for data processing and analysis. The future self-sustainability of DIAS is of critical importance for ESA. More information is expected to be released by ESA, including details around functionalities and implementation procedures for using the platform, in preparation to the launch date in June 2018.

1.3.6 CEMS[12]

32. The Climate, Environment and Monitoring from Space (CEMS) platform was established in 2013 by the Satellite Application Catapult to provide affordable and reliable cloud-computing infrastructure for the Space community, to process large volumes of data while removing the need for download and storage. The infrastructure resources are charged for individually e.g. CPU, RAM, disk and bandwidth used, and generally on a monthly basis. Unlike most cloud providers, CEMS does not charge for data transit costs i.e. ingress and egress. The SMFM team is in communication with the CEMS providers, and can arrange 1-month free trial followed by a cost estimate should the need for testing the platform arise.

1.3.7 Amazon Web Services[13]

33. This platform has a dedicated section for EO datasets[14]. The bottleneck of this platform was noted as having the file format for Sentinel-2 data that is slightly different compared to the one provided by the Sentinel Hub platform. In this case, the relevant scripts should be adjusted accordingly to be compatible with the new file format.

1.3.8 Google Earth Engine[15]

34. Google Earth Engine is a cloud-based platform for planetary-scale environmental data analysis. It combines a petabyte-scale archive of publicly available remotely sensed imagery and other data with Google's computational infrastructure optimized for parallel processing of geospatial data. This includes APIs for JavaScript and Python, and a web-based IDE for rapid prototyping and visualization of complex spatial analyses and the Landsat and Sentinel datasets. However, it is available free of charge only for research, education, and non-profit use, and accessing and processing data is not open-source, therefore not compatible with the SMFM project.

1.3.9 Google Cloud Services

35. All the Sentinel-2 data, processed at top-of-atmosphere reflectance data (Level-1C) are stored within Google Cloud Services[16]. Data processing for querying, visualising and analysing Sentinel-2 data can be done through Google Earth Engine, which is however not open-source. Google Cloud Services usage comes with a cost that covers data storage, network usage, operational usage, retrieval and early deletion fees.

1.4 Data Acquisition Plan

36. The four EO tools that are under-development as part of the SMFM project rely on a range of free and open-access data sources. Given the large scales over which the tools will operate and potentially poor access to internet, each will require a practical means of data access. The data required by the SMFM project and data access arrangements are summarised in

[11] https://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Accessing_Copernicus_data_made_easier
[12] https://sa.catapult.org.uk/facilities/cems/
[13] https://aws.amazon.com/what-is-cloud-computing/
[14] https://aws.amazon.com/earth/
[15] https://explorer.earthengine.google.com/#index
[16] https://cloud.google.com/storage/docs/public-datasets/sentinel-2

37. Table 2.

38. Some of the necessary data can be practically downloaded at national scale through web-interfaces (e.g. ALOS mosaic, land cover products) and some data will require batch-processing scripts (e.g. Sentinel-1/Sentinel-2 mosaicking). In the case of national-scale dense time-series analysis, the only practical approach may be a cloud platform where data are stored locally. In the coming sections we describe how the necessary data to operate each tool can be accessed.

Table 2 Data acquisition plan for each of the data types to be used by the SMFM project.

Data Type	Acquisition Plan
ALOS Mosaic Tiles	Data can be downloaded through a web-interface at national scales, reasonably straightforwardly. An alternative method will be provided to automate the process of download and decompression through a command line tool.
Ancillary Data (e.g. land cover products, soil moisture data)	Some data products can be improved with ancillary data (e.g. identification of flooded areas for AGB mapping). This data will be low-volume, so can be acquired through relevant websites.
Calibration and Validation Data	As either degradation or the causes of forest change cannot be readily identified in high-resolution imagery, data for tool calibration and validation will require fieldwork measurements. Fieldwork will be conducted in Mozambique and Zambia, aiming to capture a range of forest types and causes of change.
Sentinel-1 and Sentinel-2 Scenes	Data can be accessed through a graphical interface (Copernicus Open Access Hub), though this is impractical with more than 50 scenes. A command line batch download tool will be provided to assist in medium to large scale applications (e.g. national-scale mosaics). For national-scale dense time series analysis, the only practical approach may be to use a cloud platform with local access to appropriate data.

2 Tool 1a: Semi-automated Pre-processing of Sentinel-2 Data for LU/LC Classification

2.1 Objective

39. Monitoring of land cover and land cover change over large scales is only feasible with large-scale remote sensing datasets. Moderate resolution optical remote sensing instruments have been the mainstay of national-scale monitoring (e.g. Landsat), but new data from Sentinel-2 offers an increased resolution, faster re-visit times and guarantees of data continuity that make it attractive for future land cover monitoring. The SMFM Sentinel-2 mosaicking tool **'sen2mosaic'** will generate wall-to-wall mosaics at regional to national scales using composite images from multiple satellite overpasses to create season-specific cloud-free images suitable for land cover classification.

2.2 Method Design

40. The use of this tool, called sen2mosaic, includes the following steps:

- Downloading of Sentinel-2 tiles;
- Pre-processing of Sentinel-2 tiles (atmospheric correction, cloud masking);
- Generation of cloud-free composite images for each tile;
- Multi-tile mosaicking in the user-friendly GeoTiff format.

2.2.1 Downloading of Sentinel-2 Tiles

41. Image downloading is handled by a batch-processing script that accesses images from the Application Program Interface (API)[17] within the Copernicus Open Access Hub[18], which offers better performance than the web interface for large downloads. The script builds upon an existing open source Python library named sentinelsat[19]. The user is prompted to specify the tile name, date range and maximum cloud cover options to automate the selection of the most appropriate images. For the case where sen2mosaic is operated from a cloud server, where data is stored locally, this script may not be required.

2.2.2 Pre-processing of Sentinel-2 Tiles

42. Pre-processing of Sentinel-2 images is performed with the open source sen2cor toolset developed by ESA[20]. Sen2cor is a toolset to perform atmospheric correction and cloud masking of Sentinel-2 data. Sen2cor requires a level 1C (top of atmosphere reflectance) tile as input, and generates a level 2A file (atmospherically corrected, cloud masked) file as output. Sen2mosaic assists in batch processing by accepting multiple Sentinel-2 1C tiles as inputs, and automates large pre-processing tasks with parallel processing of multiple tiles simultaneously.

The cloud masks produced by sen2cor are conservative (commonly leaving in light clouds in imagery), frequently confuse cloud shadows and dark features, and can leave artefacts around image borders. A range of logical corrections using buffers and reclassification were developed and implemented to obtain an improved mosaic quality (

43. Figure 2). These are based on a modification of the methods developed by GMV (2016), used in the development of 2016 cloud-free mosaics in Mozambique. The corrections applied to the cloud mask are:

- Reclassification of 'dark feature' to 'cloud shadow';
- Reclassification of 'cloud shadow' not within 1800 metres of 'cloud' to 'water' (not masked);

[17] https://scihub.copernicus.eu/twiki/do/view/SciHubWebPortal/APIHubDescription
[18] https://scihub.copernicus.eu/
[19] https://github.com/sentinelsat
[20] http://step.esa.int/main/third-party-plugins-2/sen2cor/

- Dilation of 'cloud shadow', 'medium probability cloud' and 'high probability cloud' pixels by 180 metres to remove residual light clouds;
- Erosion of the outer 300 metres of image tiles, which removes border artefacts whilst retaining image overlap.

44. For the case where sen2mosaic is operated from a cloud server where level 2A Sentinel-2 data is stored locally, the user may opt to only run the mask correction elements of this script.

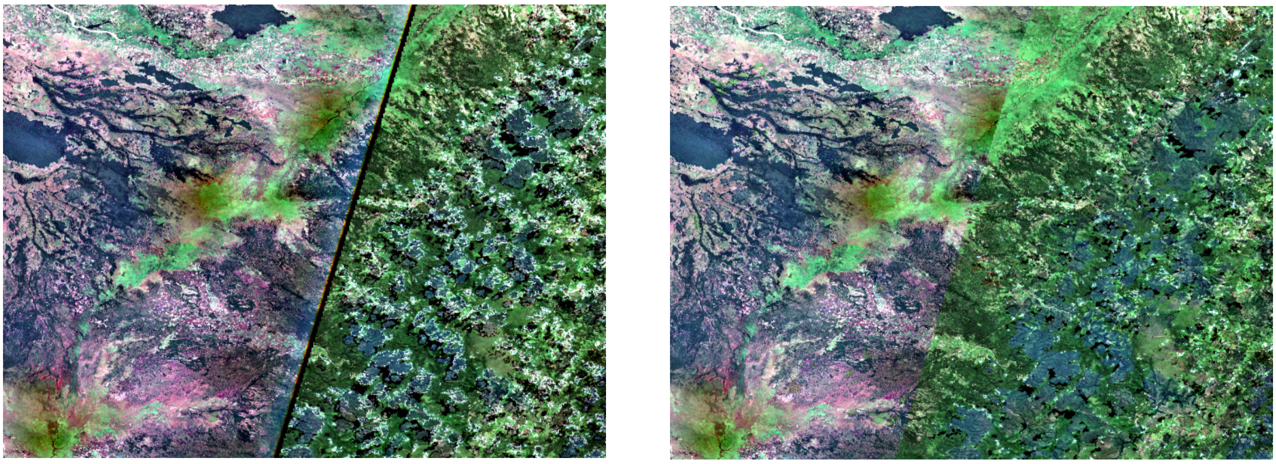

Figure 2: Mosaic outputs without modifications to the sen2cor cloud mask (left), and after modifications (right). Without modification to the cloud masks, prominent artefacts are present around light clouds and at scene boundaries.

2.2.3 Generation of Cloud-Free Composite Images

45. Atmospherically corrected Sentinel-2 tiles contain gaps where clouds have been removed from the images, and tiles that are split across the satellite swath need to be combined to create a seamless image. Sen2three is an ESA toolset for creating composite images from Sentinel-2 data, combining multiple images from each to create a single gap-filled image using available clear pixels[21].

46. Image compositing is based on the ESA tool sen2three. Sen2three is a memory intensive program, hence a series of conditions that encourage the user to input only necessary images are implemented to minimise the processing memory required.

2.2.4 Multi-tile Mosaicking in GeoTiff Format

47. The final step generates a seamless mosaic with composite images re-projected to a single coordinate reference system, including the conversion to the user friendly and flexible GeoTiff format. The script also outputs VRT files, which are virtual raster file that can be used for straightforward visualisation in open source GIS software such as QGIS. This script uses the Geospatial Data Abstraction Library (GDAL)[22], a widely used open source Python library for dealing with geospatial data.

2.3 User Interface and Usage

48. Sen2mosaic will be made available as a set of four scripts to perform the processes of downloading, pre-processing, compositing and mosaicking. These scripts will be usable from the Linux command line, or by a more advanced user by importing functions in Python.

National-scale outputs from sen2mosaic are shown in Figure 3.

---

[21] http://step.esa.int/main/third-party-plugins-2/sen2three/

[22] http://www.gdal.org/

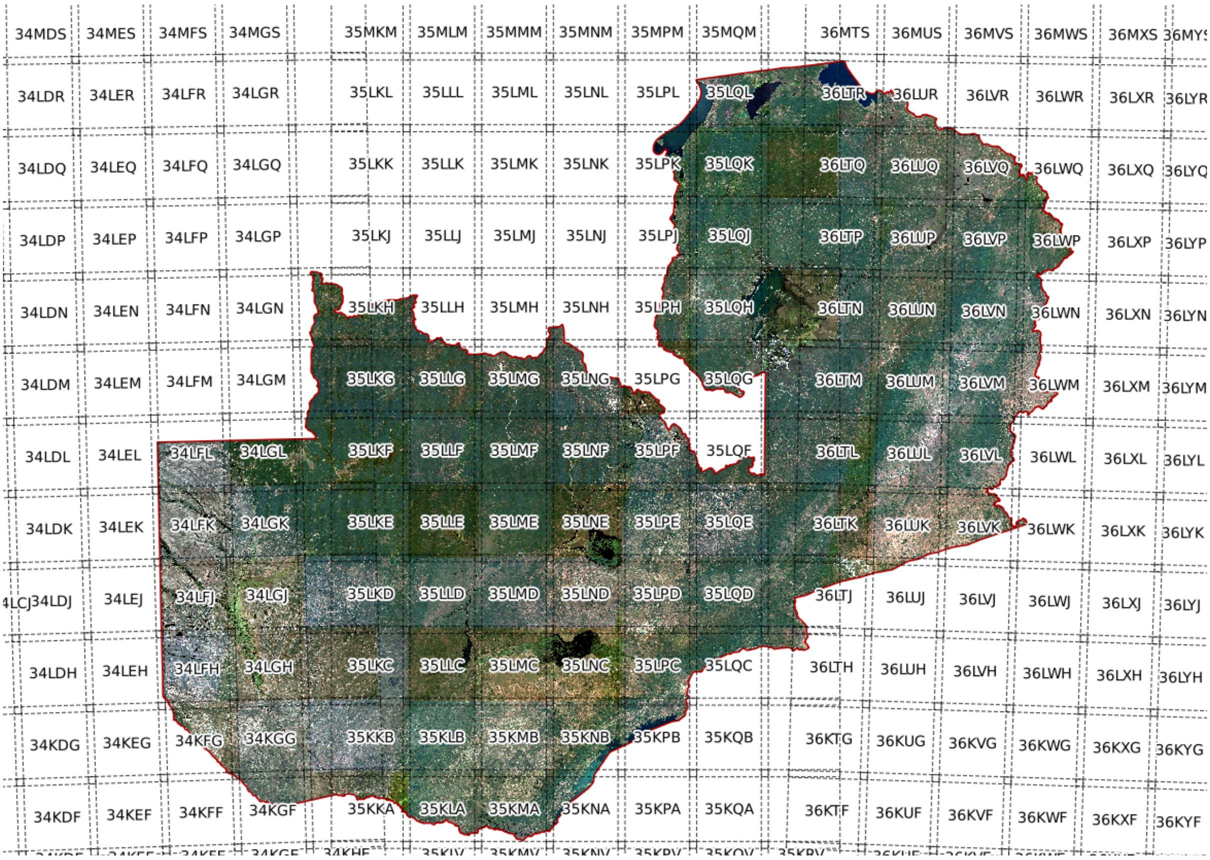

Figure 3: A 10 metres resolution mosaic of Sentinel-2 data generated by sen2mosaic for the extent of Zambia, using images for the early dry season (May – June) 2017.

2.4 Documentation and Distribution

49. The SMFM sen2mosaic scripts are publicly available on bitbucket[23]. Users must also install ESA tools sen2cor and sen2three, as well as the open source sentinelsat tool from github[24].

Documentation and worked examples are available online[25].

[23] https://bitbucket.org/sambowers/sen2mosaic
[24] https://github.com/sentinelsat
[25] http://sen2mosaic.readthedocs.io/en/latest/

3 Tool 1b: Semi-Automated Pre-processing of Sentinel-1 Data for LU/LC Classification

3.1 Objective

50. Land cover mapping most commonly makes use of optical data, which is widely available and comparatively easy to work with. Using additional data from radar instruments can provide useful and complementary inputs to land cover classifications, but uptake has been limited by the limited availability of free data and onerous pre-processing requirements. With the launch of the Sentinel-1 radar sensors with free and open-access data there is an opportunity to leverage new datasets to improve land cover classifications. The SMFM Sentinel-1 mosaicking tool **'sen1mosaic'** will assist in the generation of large-scale Sentinel-1 mosaic products suitable for use in land cover classification tasks.

51. During the inception phase, it was agreed to test the utility of Sentinel-1 data for improving land cover classifications, and develop a fully functional tool if the outputs were judged to be useful (SMFM Inception Report, Section 6.2.1). In practice, it was found that the burden of testing on partner countries would be difficult to manage and that SMFM Tool 3 will likely require pre-processing of Sentinel-1 data anyway, therefore the complete tool to pre-process Sentinel-1 data (i.e. Tool 1b) has been developed in advance of a testing phase.

3.2 Method Design

52. Operation of this tool, named sen1mosaic, takes the following steps:

- Downloading of Sentinel-1 images;
- Pre-processing of Sentinel-1 images;
- Creation of multi-image mosaics in the user-friendly GeoTiff format.

3.2.1 Downloading Sentinel-1 Images

53. Like sen2mosaic, sen1mosaic uses the sentinelsat library to download images. Sen1mosaic uses Ground Range Detected (GRD) images collected in the Interferometric Wide (IW) swath mode, the main acquisition mode of Sentinel-1 and the most applicable to land use studies. Data from Sentinel-1 are not provided in a standard tiling grid like Sentinel-2, and images are acquired from both ascending and descending paths. The tool requires the user to input a bounding box in latitude/longitude, and downloads all images that meet criteria of date ranges and specification of ascending or descending paths. For the case where sen1mosaic is operated from a cloud server, where data is stored locally, this script may not be required.

3.2.2 Pre-processing of Sentinel-1 Images

54. Image pre-processing is performed using the Sentinel Application Platform (SNAP) graph processing tool[26]. SNAP is an open source ESA program which has functions for pre-processing and analysis of satellite imagery, including a range of Sentinel-1 specific functions.

55. Image pre-processing takes the following steps using SNAP functions:

- **Application of orbit file**, which improves image geo-referencing through precise orbit state data.
- **Thermal noise removal** applies corrections for thermal noise, which is particularly problematic in the cross-polarised (VH) channel (optional)
- **Calibration** from digital numbers to pixel values directly related to radar backscatter in the scene.
- **Slice assembly** stitches together consecutive satellite scenes to produce a continuous image for each satellite pass.
- **Multi-looking** (optional) reduces noise and processing load by degrading image resolution.

[26] http://step.esa.int/main/toolboxes/snap/

- **Radiometric terrain flattening** (optional) corrects for the tendency for slopes facing towards the sensor appearing brighter than those facing away.
- **Filtering** (optional) further reduces image noise from speckle. We apply a Refined Lee filter, a radar-specific speckle filter which aims to preserve edges between field boundaries in images.
- **Geometric terrain correction** aligns pixels with their geographical location, and corrects for the tendency of topographic features to appear to lean towards the sensor (foreshortening).
- **Translation and mosaicking** aligns multiple radar passes to a common Universal Transverse Mercator (UTM) geographical grid. We output processed images in units of Decibels.

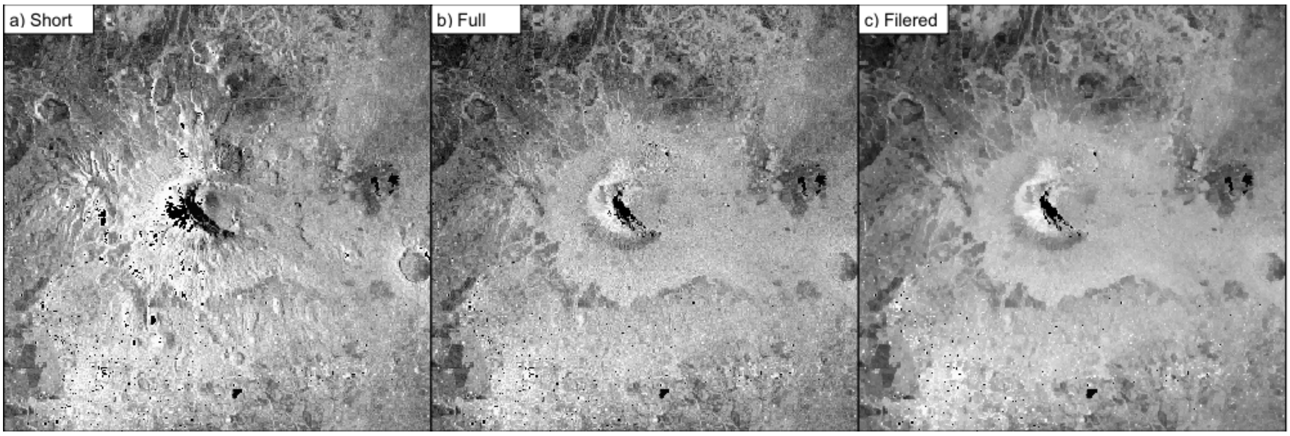

Figure 4: Example outputs from sen1mosaic (VV backscatter) showing processing options of (a) short processing chain, (b) full processing chain and (c) full processing chain with speckle filtering. This example shows Mount Meru in Northern Tanzania.

56. A number of processing steps in sen1mosaic are optional. Speckle filtering can be activated or deactivated to accommodate multiple user's needs, and the multi-looking step will be customisable to allow specification of output resolutions. As this processing chain is computationally intensive, a separate short processing chain (omitting thermal noise removal and radiometric terrain flattening) can be optionally run to generate test outputs rapidly. A range of outputs using different processing chains are shown in Figure 4.

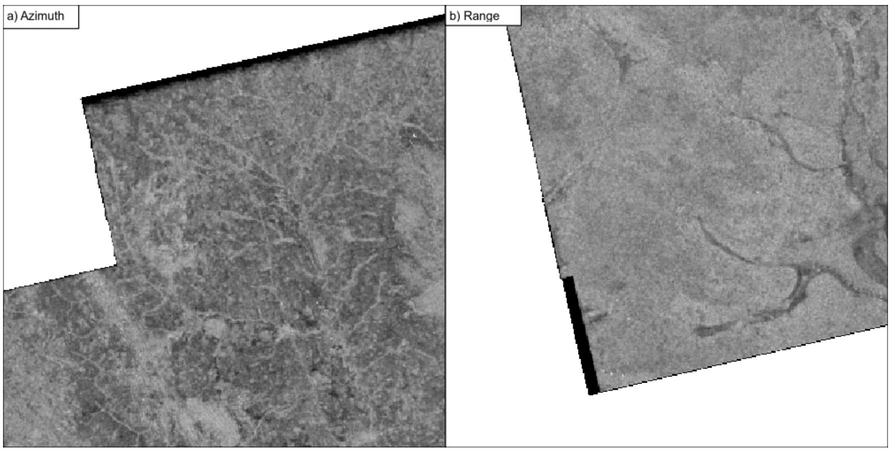

Figure 5: Sentinel-1 border artefacts from the (a) azimuth (along track) and (b) range (across track) directions, which are removed by sen1mosaic.

Sentinel-1 data outputs processed in SNAP are prone to image artefacts that can interfere with image classification (

57. Figure 5). Image sub-setting is used to eliminate border noise and minimise along-track artefacts whilst retaining the overlap between satellite overpasses.

For an advanced user, the tool can also support the use of customised processing chains complied in SNAP.

3.2.3 Multi-image Mosaicking in GeoTiff Format

58. A mosaicking script very similar to that used by sen2mosaic (Section 2.2) is used. The main difference is that in place of the compositing and mosaicking steps of sen2mosaic, sen1mosaic generates outputs of the mean average, minimum, maximum, and standard deviation of each pixel in input images, which together describe the backscatter properties and its temporal variability of each pixel which can be used as inputs in land cover classifications. The output is provided in GeoTiff format, in the UTM zone and resolution specified. Two additional VRT files are provided for rapid visualisation of outputs in open source GIS software (e.g. Figure 6).

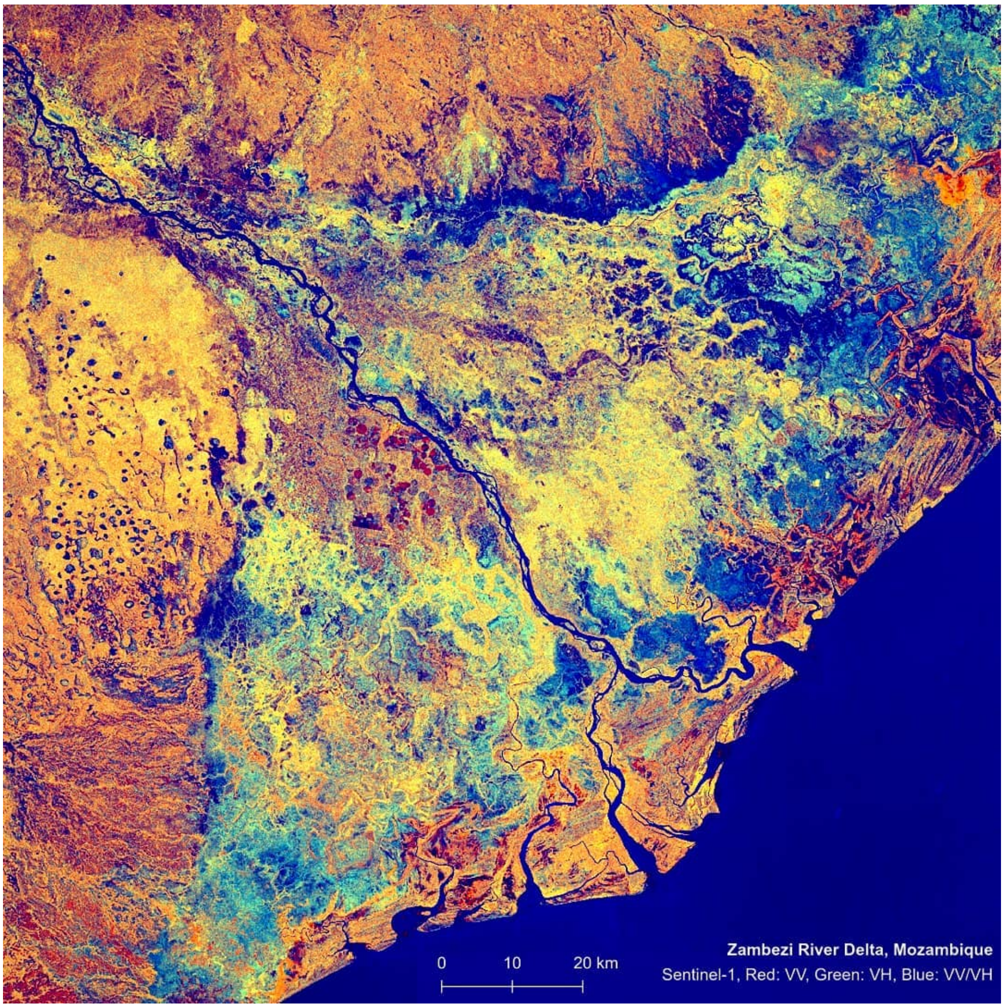

Figure 6: Highlight from a regional-scale mosaic of Sentinel-1 data for the period May-June 2017 for Zambezia Province of Mozambique, generated with SMFM sen1mosaic.

3.3 User Interface and Usage

59. Sen1mosaic will be made available as a set of three scripts to perform the processes of downloading, pre-processing, and mosaicking. These scripts will be usable from the Linux command line, or by a more advanced user by importing functions in Python. The scripts required to operate sen1mosaic will mirror sen2mosaic, and will therefore require similar user expertise to be executed.

3.4 Documentation and Distribution

60. The SMFM sen1mosaic scripts are publicly available on bitbucket[27], with documentation and worked examples are also available online[28].

[27] https://bitbucket.org/sambowers/sen1mosaic
[28] http://sen1mosaic.readthedocs.io/en/latest/

4 Tool 2: Annual Forest Biomass Change and Degradation Mapping Using the ALOS PALSAR Mosaic

4.1 Objective

61. Long wavelength radar data are a long-established method of generating maps of forest AGB in woodlands and dry forests (Le Toan et al. 1992, Le Toan et al. 2011, Ryan et al. 2012, McNicol et al. *in press*). Free data are now available from JAXA's ALOS-1 and ALOS-2 L-band sensors (2007 – 2010, 2015 – onwards) (Shimada et al. 2010, Motohka et al, 2012), which offer a relatively straightforward means to generate maps of AGB in woodlands and forests with AGB of up to around 150 tonnes per hectare. The SMFM **'BIOmass'** Tool for ALOS (biota) is a tool that generates maps of AGB, AGB changes between multiple years, and uses estimates of AGB change to classify locations of deforestation and degradation.

4.2 Method Design

62. Use of the Biomass tool has four elements:

- Downloading ALOS mosaic tiles;
- Pre-processing ALOS data;
- Mapping aboveground biomass and forest cover;
- Detection of forest changes.

4.2.1 Downloading ALOS Mosaic Tiles

63. Data from the ALOS mosaic can be accessed from JAXA through a graphical interface online after signing up[29]. The data is delivered in either 1x1 degree tiles or 5x5 degree collections of tiles. To obtain national-scale data using this interface is possible, but for large-scale applications a command-line interface to automate data download is considered preferable.

64. The biota tool includes a script to download ALOS mosaic tiles directly from the JAXA FTP server. The user specifies either a 1x1 or 5x5 degree tile by its upper-left corner latitude and longitude and a year or set of years. The downloader handles access of both the older ALOS-1 filename structure and the ALOS-2 filename structure. The script also optionally decompress images after download.

4.2.2 Pre-processing ALOS Mosaic Data

65. The ALOS mosaic is provided as a series of 1x1 tiles containing composites of multiple overpasses by ALOS-1 and ALOS-2. The format of the ALOS mosaic data is described in detail in its user manual[30]. Backscatter data is stored as 16-bit unsigned digital numbers for each of HH and HV polarisations, and require the application of a calibration equation to generate gamma nought (gamma0) backscatter values in units of decibels:

$$\gamma_0 = 10log_{10}(DN^2) - 83.0\ dB$$

66. Biota is a Python module built to automate pre-processing tasks of large volume of data over large areas in a programming environment. The steps for calibration of ALOS mosaic data to a usable object in Python are:

- Loading of data into memory, and extraction of metadata from filenames.

[29] http://www.eorc.jaxa.jp/ALOS/en/palsar_fnf/fnf_index.htm
[30] http://www.eorc.jaxa.jp/ALOS/en/palsar_fnf/DatasetDescription_PALSAR2_Mosaic_FNF_revE.pdf

- Conversion and calibration to gamma0 backscatter.
- Application of “bad data” mask, with additional optional masking of data from the wet season and around water bodies and flooded areas.
- Optional filtering to reduce image noise with an Enhanced Lee Filter, a radar speckle filter that aims to preserve edges.
- For more advanced users, the tool supports the output of gamma0 data to Geotiffs for customised onward processing chains.

67. In the pre-processing of radar data, order of processing steps is important. The biota module enforces the use of pre-processing steps in an appropriate order, and through its structure discourages inappropriate use of functions.

4.2.3 Mapping Aboveground Biomass and Forest Cover

68. The module includes a generic equation to calibrate gamma0 backscatter to forest AGB (McNicol et al. *in press*). This equation was developed for southern African woodlands using forest plot data from Mozambique, Tanzania and Malawi, using ALOS-1 data. This model for southern Africa can be applied as a first estimate of AGB where appropriate forest plot data does not yet exist.

69. For global applicability, the tool supports the calibration of country-specific backscatter-AGB relationships using available national forest plot data. Functionality will be provided to ingest shapefiles containing forest plot data and generate a linear or logarithmic calibration equation to relate radar backscatter to AGB. This is of particular importance under the SMFM project as one of the objectives is to provide generic and globally applicable methods and tools.

70. The biota tool generates forest maps that match and can be tailored to nationally-determined parameters, through the following variables:

- A forest AGB threshold (in tonnes per hectare) to separate forest from non-forest;
- A minimum area threshold (in hectares) of contiguous forest.

Maps of AGB and forest cover can be output as GeoTiffs for straightforward visualisation and generation of summary statistics (Figure 7 a/b).

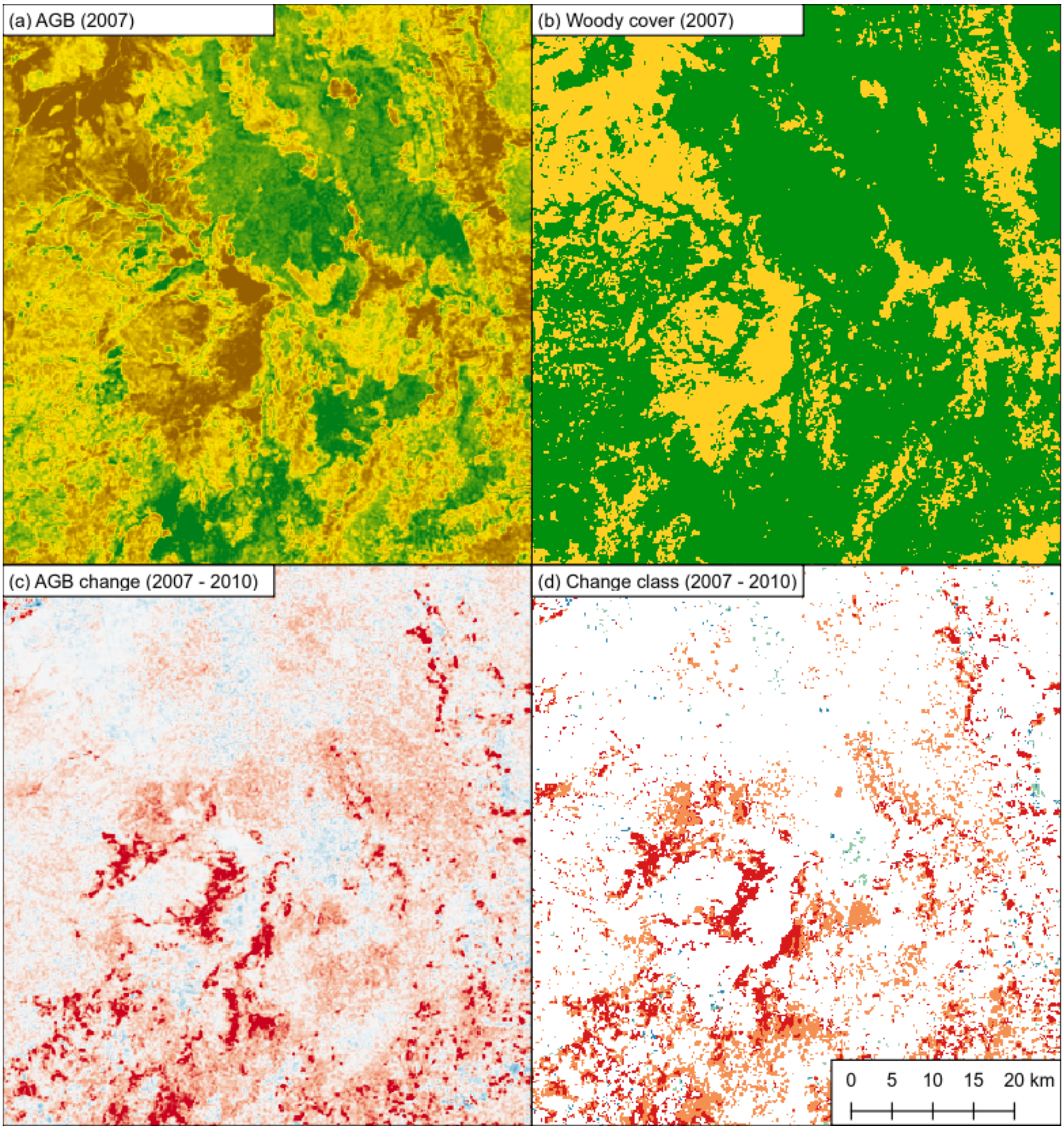

Figure 7: Maps of (a) aboveground biomass (AGB) (brown = low, yellow = moderate, green = high), (b) forest cover (yellow = non-forest, green = forest), (c) AGB change (red = loss, blue = gain) and (d) change classification (red = deforestation, orange = degradation, blue = growth), generated by the biota tool. Maps show a location in Kilwa District of south eastern Tanzania, which underwent rapid land-use change from the expansion of smallholder agriculture between the years 2007 to 2010. These maps are based on data from ALOS-1, a generic southern African biomass-backscatter relationship and an arbitrary definition of forest.

4.2.4 Change Detection

71. "False positives" in the detection of forest change are a common problem to using data from ALOS because of the noise inherent to radar images. This issue is compounded in use of data from the ALOS mosaic, which may not come from the dry season and cannot be averaged out over multiple images. Thus, a naive classification of a difference image based on AGB change thresholds alone will tend to overestimate rates of forest change. The following strategies are employed to suppress noise in radar imagery and to reduce its impact on change predictions were implemented:

- **Filtering**: The Enhanced Lee filter suppresses noise. Applied to the gamma0 images in the pre-processing steps.

- **Minimum change area specification:** In addition to matching a national forest definition, minimum areas of change can be employed to remove isolated pixels of apparent change.

- **Minimum change intensity thresholds:** Spurious changes may be recorded where changes of small magnitude cross the threshold of forest/non-forest. A minimum intensity of change (e.g. 20% loss of AGB) can be used (optional) to filter out some of the changes that likely result from noise in data.

- **Masking of water**: In addition to AGB, L-band radar backscatter is sensitive to soil moisture. Areas with seasonal/inter-annual soil moisture changes may therefore be interpreted as changes to AGB. A buffer of user-defined size around rivers and lakes was applied to remove all changes that occur near rivers and water bodies. Ancillary land cover maps can be used to identify areas of frequent flooding to also be masked.

- **Soil moisture correction:** Inter-annual variation in backscatter resulting from varying soil moisture may also be corrected using soil moisture estimates from other remote sensing products. We will investigate the utility of these corrections further for the SMFM biota tool.

72. Biota identifies areas of forest change by identifying locations that change between two years of measurement (e.g. 2007-2010). By default, the tool identifies six classes of change (Table 3), which can be modified to suit national definitions of forest over and forest cover change. In addition to the area of change, the biota tool outputs the magnitude of AGB that can be attributed to each change event. Example outputs are shown in Figure 7.

Table 3: Default definitions of forest change in biota. Changes are classified based on a forest AGB threshold (forest as opposed to non-forest), an intensity threshold (minimum proportional change of AGB) and a minimum area threshold (minimum size for a change event). Source: modified after McNicol et al (2018)

Forest Change Type	Definition
Deforestation	AGB above forest threshold in image 1 and below forest threshold in image 2, with a loss of AGB greater than intensity threshold over an area greater than minimum area threshold.
Degradation	AGB above forest threshold in image 1 and image 2, with a loss of AGB greater than intensity threshold over an area greater than minimum area threshold.
Minor loss	Any loss of AGB that doesn't meet conditions of deforestation or degradation*.
Minor gain	Any increase of AGB that doesn't meet conditions of growth or afforestation*.
Growth	AGB above forest threshold in image 1 and image 2, with an increase of AGB greater than intensity threshold over an area greater than minimum area threshold.
Afforestation	AGB below forest threshold in image 1 and above forest threshold in image 2, with an increase of AGB greater than intensity threshold over an area greater than minimum area threshold.
Non-forest	AGB below forest threshold in image 1 and image 2

**This is required for carbon balance but may not be important for forest mapping and monitoring purposes.*

4.3 User Interface and Usage

73. The biota tool is provided as a Python module containing a range of functions to load, calibrate, produce maps of AGB and classify forest change, and generate summary statistics using ALOS mosaic tiles.

74. The Python scripts required to use this tool should be straightforward to implement even by users with limited programming experience. Although the module is still under development, an indicative script for producing two biomass maps and a change map for a 1x1 degree tile is shown in Figure 8.

```
import biota

# Load data for 2007/2010
tile_2007 = biota.LoadTile('~/DATA/', -18, 33, 2007, lee_filter = True, forest_threshold = 15,
area_threshold = 2.5)
tile_2010 = biota.LoadTile('~/DATA/', -18, 33, 2010, lee_filter = True, forest_threshold = 15,
area_threshold = 2.5)

# Generate AGB maps and output GeoTiff files
AGB_t1 = tile_2007.getAGB(output = True)
AGB_t2 = tile_2010.getAGB(output = True)

# Generate a forest cover maps for 2007
forestcover_2007 = tile_2007.getForestCover(output = True)

# Add buffer around riverine areas to reduce false positives from soil moisture
tile_2010.updateMask('rivers.shp', buffer_size = 750)

# Load a change object
tile_change = biota.LoadChange(tile_2007, tile_2010, change_intensity_threshold = 0.25,
change_area_threshold = 2.5)

# Generate a deforestation/degradation map and output to GeoTiff
tile_change.CalculateChange(output = True)
```

Figure 8: Example of Python script for generating a maps of AGB (2007 and 2010), a forest cover map (2007) and a forest change map (2007 to 2010) for the ALOS tile with latitude -18 and longitude 33. An arbitrary threshold of 15 tonnes of carbon per hectare over an area of at least 2.5 hectare is used to distinguish forest, and changes must change AGB by at least 25% and over an area of at least 2.5 hectare.

4.4 Documentation and Distribution

75. The (under development) biota module is available online[31], with documentation, which will include a number of worked examples for common processing tasks, in progress[32].

[31] https://bitbucket.org/sambowers/biota
[32] http://biota.readthedocs.io/en/latest/

5 Tool 3: Continuous Forest Change Monitoring with Sentinel-2 Data

5.1 Objective

76. The reliable measurement of deforestation in tropical forests is an ongoing challenge, and remote sensing of forest degradation is still in a research phase. These challenges are particularly great in tropical dry forests, which have strong seasonal cycles, are highly heterogeneous and show large inter-annual variations. New data from the Sentinel-1 and Sentinel-2 platforms offer unprecedented resolution and frequency for freely available data, and may be capable of reducing uncertainty in rates of deforestation and degradation in tropical dry forests.

77. The SMFM DEnse FORESt Time series tool **'deforest'** will aim to use data from Sentinel-1 and Sentinel-2 to overcome some of the monitoring issues of tropical dry forests through analysis of dense time series of data. The tool will aim to be both useful for accounting of historical forest changes along with near real-time monitoring of forest change for management purposes.

5.1.1 Research Plan

78. Approaches for extracting information on forest change from dense time series are still experimental, with a range of methods that have been proposed. Three major challenges exist in classification of dense time-series of imagery from dry forests and woodlands:

- Cloud cover is a problem in optical data, especially during the wet season;
- Inter and intra annual variability in rainfall, tree-leaf display and grass growth leads to significant variations in land surface phenology, as does the occurrence of fire. Such changes can often create false positives in change detection algorithms
- Dry forests are characterised by landscape heterogeneity, which means that the signal associated with deforestation and degradation is likely to differ across a landscape.

79. A recent paper by Reiche et al. (2018) offers some potential solutions to these challenges through multi-sensor integration and image normalisation. Reiche et al. (2018) used a multi-sensor dense time series of data (Landsat, Sentinel-1 and ALOS PALSAR) to generate near real-time estimates of deforestation in Bolivia. This method uses a Bayesian framework to flag forest changes to generate early warnings of forest change, and confirms them over time as more data becomes available. This method offers an elegant means of combining long term monitoring with shorter term alerts that can be actioned by forest managers.

80. For the purposes of SMFM project, this method has a number of advantages. Foremost is that there is no requirement for long time series of data to calibrate the model, a desirable property where using data from recently launched satellite sensors. Limitations of cloud cover are addressed through integration of multiple sensors, and the method includes an approach for removing the effects of seasonality from imagery ('de-seasonalisation') which is not reliant on prior knowledge of the seasonal cycle. Finally, the data processing and memory requirements are greatly reduced relative to time-series approaches such as BFAST, requiring each image to be loaded into memory only once, rather than building a profile of each pixel through time.

81. Firstly, the development of the tool will focus on the reliable detection of deforestation events, which, being of greater magnitude is a comparatively straightforward problem to solve. Secondly, following the collection of field data identifying locations of degradation (fieldwork planned for May-July 2018), the tool will be adapted to classify degradation events. Even if the higher risk goal of detecting forest degradation does not prove reliable, the tool will still provide novel utility that will be useful to the tropical dry forest monitoring community.

5.2 Method Design

82. The prototype deforest tool has three steps:

- Data download and pre-processing;

- Image classification;
- Bayesian detection of change events.

5.2.1 Data Download and Pre-processing

83. The download and pre-processing steps for the deforest tool will use the SMFM sen2mosaic and sen1mosaic tools (section 2 and section 3, respectively). The cloud masking and artefact removal algorithms within these tools are particularly important as extreme pixel values may be interpreted as deforestation event. In terms of data processing and storage this is the most computationally intensive processing step, with the volumes of data required by the deforest tool necessarily large (approximately 1 terabytes for 100 x 100 km over 3 years).

5.2.2 Image Classification

84. The Bayesian change detection algorithm requires that all input images are converted into maps of the probability of forest (or degraded forest) being present. This transformation of images from multiple sensors into the same units allows them to be compared, and for progressive certainty for the detection of forest changes.

85. Classifiers of forest cover will vary widely, and in the same way that different countries will be unlikely to use the same classifier for land cover mapping, solutions will likely differ for forest probability estimation. Therefore, the scripts for calibration were separated from those for change detection, which will be usable independently of the classification method. Here is described the classification approach for converting input images into probabilistic maps of forest.

86. **Feature Selection -** Optical sensors have multiple inputs bands which contain information about land cover. For example, in optical data (e.g. Sentinel-2) the red and near-infrared spectral bands contain information on the green vegetation components. Spectral bands can be combined into spectral indices that highlight the property of interest (e.g. NDVI), in this case the presence or condition of vegetation cover. As well as highlighting vegetation, spectral indices have the property of mitigating against differences in scene illumination or topography.

87. Radar sensors do not have an equivalent of spectral bands, but many are equipped to transmit and measure radiation with different polarisation (orientation). For radar imagery (e.g. Sentinel-1) the tool uses input features of backscatter in each available polarisation and ratios between them.

88. **Image normalisation -** The SMFM deforest tool applies an image normalisation technique called 'de-seasonalisation' (Hamunyela et al. 2016) to input images, which aims to normalise for seasonal effects in imagery. In its original form this involves running a moving window over the image and taking the 95th percentile of an index (e.g. NDVI), and dividing by this value. The rationale is that this value should always represent a forest area (given a large enough window), and therefore is indicative of the phenological state of forests at that location. Reiche et al. (2018) used a simplified version of this method where, instead of a moving window, the 95th percentile of a baseline map of forest for each input image was taken, and subtracted this value from other pixels in the image.

89. De-seasonalisation is potentially powerful as it can be applied to images at any time of year without a long calibration period, it is robust to drought years, and with a moving window it should take into account landscape heterogeneity. Both the moving window approach of Hamunyela et al. (2016) and the forest baseline method of Reiche et al. (2018) will be implemented in the deforest tool, allowing their outputs to be tested and compared.

90. **Logistic Regression -** Logistic regression is used in the deforest tool to convert input features into a probability of forest cover. Logistic regression is an appropriate classifier as it returns well-calibrated probability estimates[33], and where input features are correlated the model can be made robust to overfitting through regularisation (which prioritises a 'simple' model to a more complex one). A separate model must be fit for each input sensor (Sentinel-1 and Sentinel-2), and different models for each acquisition mode (e.g. Sentinel-1 single polarisation, Sentinel-1 dual polarisation).

[33] http://scikit-learn.org/stable/modules/calibration.html

91. Each model will be calibrated with data from training regions representing forest and non-forest areas (or degraded and non-degraded) locations, with model predictions tested using a cross-validation process. The resultant model parameters are used to classify all input images, outputting an ordered series of GeoTiff files showing the probability of forest at each location in each image.

5.2.3 Bayesian Detection of Change Events

92. The Bayesian change detection approach of Reiche et al. (2018) is used by the deforest tool. The method takes a time-series of forest probabilities, combining new information on forest state with previous observations to generate updated probability estimates of forest being present in a pixel at each point in the time series (P_{forest}). Where an observation with < 50% probability of being a forest is encountered, a potential deforestation event is flagged. P_{forest} is updated by future observations, leading to the eventual acceptance of the event where P_{forest} passes a user-specified threshold (depending on the degree of confidence required), or rejection of the change as a false positive where P_{forest} falls back below 50%. Observations are subject to a 'block-weighting function', which limits the probability of forest in an individual observation to the range 10% to 90%, preventing extreme observations from resulting in false detections of forest change.

93. The original module (available on github[34]) is written in R, but by its design as an experimental system is computationally inefficient. The deforest tool improves the efficiency of the original script significantly by processing entire arrays simultaneously in place of the original pixel-by-pixel approach.

Outputs of the Bayesian change detection algorithm from the prototype scripts for monitoring and early-warning are shown in Figure 9 and

94. Figure 10.

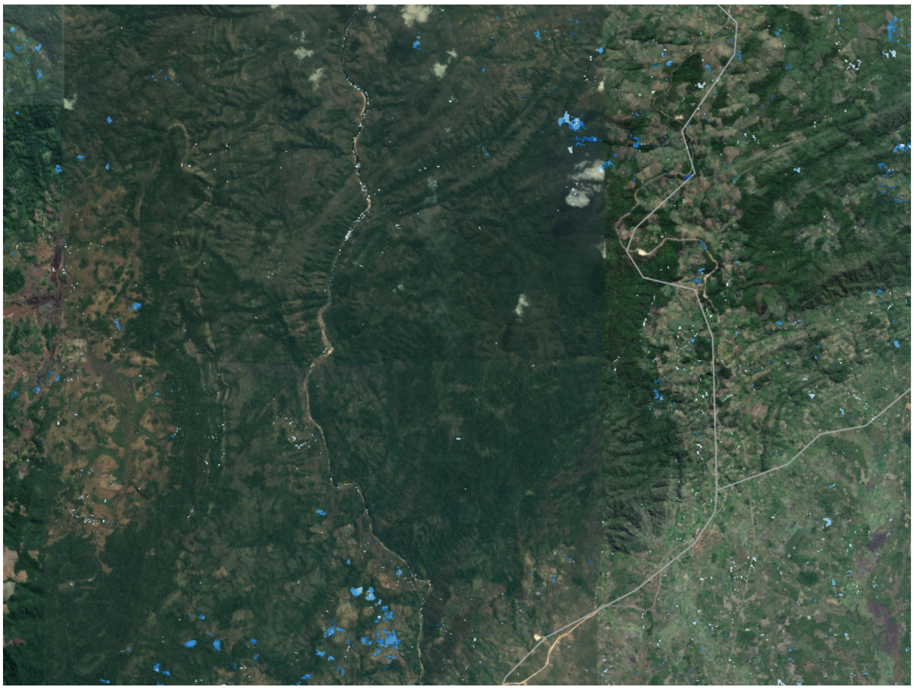

Figure 9: Outputs from the prototype deforest tool, showing 'confirmed' deforestation events in blue for the year 2017 in the vicinity of Moribane forest reserve in Central Mozambique, one of the locations selected to perform fieldwork.

[34] https://github.com/jreiche/bayts

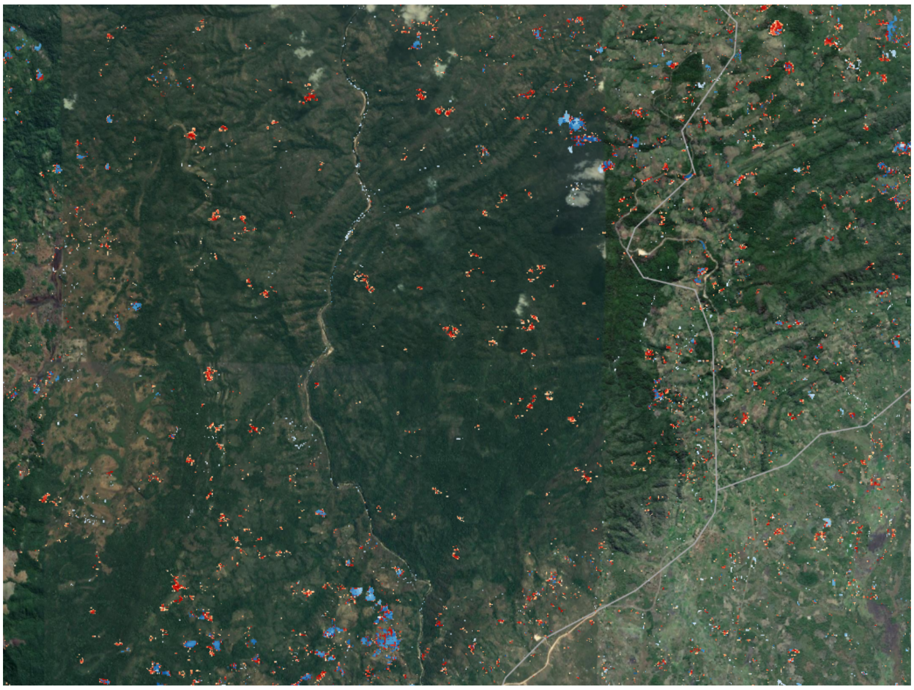

Figure 10: Outputs from the prototype deforest tool, showing 'early warning' deforestation events in red for the year 2017. Many of these events will later be rejected as false positives, but this near real-time view is useful as a management tool.

5.3 User Interface and Usage

95. The deforest tool is provided as a Python module with a command-line interface. The command line tools can be operated in a similar fashion to those of the SMFM sen2mosaic and sen1mosaic tools.

96. Like image classification for land cover mapping, users will require a country-specific classification algorithm, which cannot be generalised. Therefore, some knowledge and guidance is required to calibrate the tool for an individual location, whereas ongoing monitoring should be more straightforward once tools are calibrated.

5.4 Documentation and Distribution

97. The scripts for the deforest tool (under development) are available on bitbucket[35], and will include related documentation (not yet provided)[36].

[35] https://bitbucket.org/sambowers/deforest
[36] http://deforest.readthedocs.io/en/latest/

6 Tool 4: Identifying Causes of Forest Change

6.1 Objective

98. In addition to identification of the locations of land cover change, managers of dry forests may also be interested in identification of the causes and drivers of land cover change. Attributes such as the size, shape and intensity of disturbance events may provide clues about the type of change, which may be used to classify change events by their cause. The aim of this tool is to identify the cause to individual forest cover change events detected by remote sensing analysis.

99. Classification of forest change type is experimental and has not yet been attempted in tropical dry forests. Here, a novel method to identify the cause of forest change events in remote sensing data products is outlined. It is noted that the novel and innovative nature of this research comes with some risk and uncertainty, with very little similar existing work available the SMFM team.

This tool will be developed after field data collection has been conducted.

6.2 Method Design

100. This tool will have three components:

- Fieldwork data collection;
- Training a classifier and validation;
- Prediction of change events.

6.2.1 Fieldwork Data Collection

101. This tool will require data from the field to calibrate and validate its predictions. It is common to use data from Google Earth to calibrate and validate land cover / land cover change maps (e.g. Collect Earth tool), but as it is not possible to determine the cause of forest changes or to reliably identify degradation even using high-resolution imagery, the tool will always require local field data to be calibrated.

102. Given the effort and cost associated with field data collection, the field data collected should be usable for multiple purposes. The objectives of field data collection (May-July 2018) will therefore be (in order of importance):

- **Objective 1**: Obtain reference data that identifies locations of woody loss (deforestation and degradation), and their causes (e.g. logging, charcoal, agriculture etc.) to provide training data;
- **Objective 2**: Collect validation data to test the Producer's accuracy of the deforestation/degradation predictions of tool 2 and tool 3 (i.e. the probability of an event on the ground being detected by the remote sensing product);
- **Objective 3**: Collect validation data to test the User's accuracy of the deforestation/degradation predictions of tool 2 and tool 3 (i.e. the probability of an event identified by the remote sensing product being present on the ground).

103. The plan will involve stratifying the landscape into representative land cover types (e.g. tropical dry forest and woodlands), sampling to test and compare the tool's performance in both land cover types. The main field activity will be searching for areas of recent (2017 or later) woody cover change. Upon identifying a change event, its extent will be recorded on a GPS unit, pictures will be taken, its cause will be identified and documented through consultation of local experts, and the time at which it occurred will be estimated. A number of change event measurements will be necessary for each type of change (n = 30), aiming to include as many different causes as possible. For some causes this will be relatively straightforward (e.g. smallholder agriculture), and others likely much harder (e.g. logging). This data will be used for objectives 1 and 2. Objective 3 is more problematic, as it should only be performed once the tools are finalised. Additionally it is

much harder to travel to a specific event than to find one of many examples of forest change. For objective 3, methods will be trialled with the aim of informing future fieldwork efforts.

6.2.2 Training a Classifier and Validation

104. GPS tracks will be used to match change events in remote sensing datasets to their cause. The tool will extract a range of features from each change events (e.g. size, shape, intensity etc.), which will be used to train a classifier to identify the cause of change events in satellite data (Figure 11). A range of classification approaches will be tested, including machine learning techniques, with the eventual choice of classifier will depend on the field data and through quality assessment of the classifier. The algorithm will be validated as a cross-validation, where a proportion of the field data are held back for validation purposes.

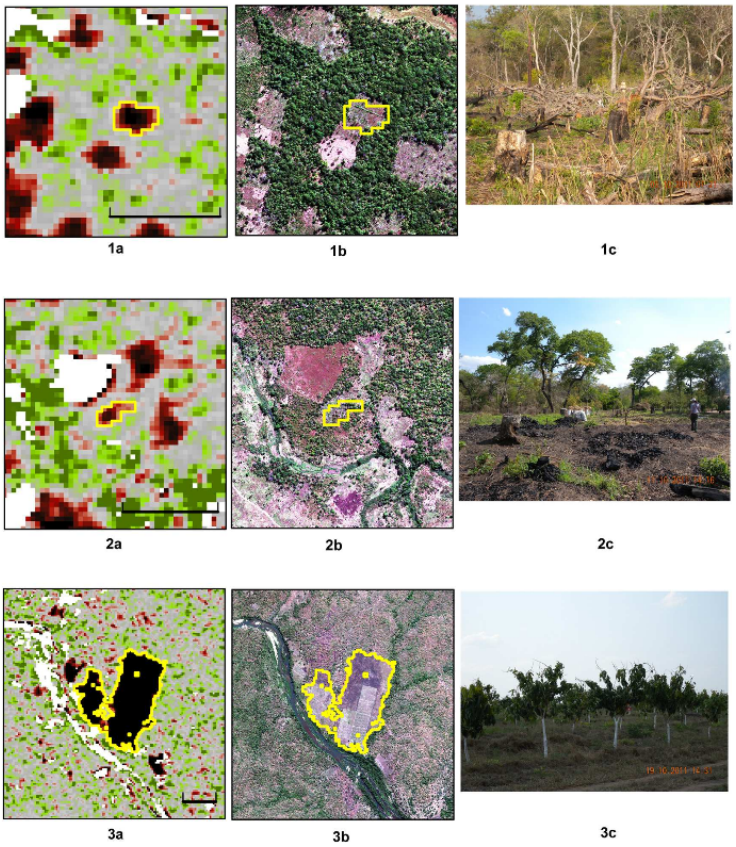

Figure 11: Examples of individual change events identified using ALOS PALSAR data in Mozambique (Ryan et al. 2014), which will be used to classify change types in remote sensing imagery.

6.2.3 Prediction of Change Events

105. The tool will use functions within Python to extract individual change events based on criteria of minimum extent and intensity. Functions will extract key features from each change (e.g. size, shape etc.), which will be input into the classification algorithm to identify its cause. The tool will output GeoTiff files for visualisation in addition to reports of the area attributable to each change. The classifier will be applied to the outputs of tool 2 and tool 3, but should be equally applicable to other commonly used remote sensing datasets (e.g. Hansen et al. 2013).

6.3 User Interface and Usage

106. The tool will be provided as a series of Python scripts operated from the Linux command line. For more advanced users there will be a Python interface, and detailed methodological documentation to assist in further development of the method in other programming languages (e.g. R).

6.4 Documentation and Distribution

107. Documentation will follow finalisation of the tool.

7 Tool Development Plan

There are key planned steps within the project program at which the tools will have advanced to a point where they can be tested, and associated training activities implemented. Thereafter, the tools will be useable and considered completed, but, with the nature of developing open-sourced tools and products, the tools will be developed and adjusted throughout the duration of the project. This approach will allow changes to be made based on user feedback ('interactive development'), and adjustments to be made to approaches with availability of field validation data. There is some uncertainty as to the timing of implementation, given developments to the DIAS platform on which the tools will be hosted, as the functionality and release date of this platform is uncertain. The current development plan, outlining key dates and steps in relation to the SMFM tools can be seen below in Table 4.

Table 4. Tool Development Plan for SMFM Project

SMFM Satellite EO Method Design

Tool No.	Description	Year / Month / No.	J	F	M	A	M	J	J	A	S	O	N	D	J	F
		Year	2018												2019	
		No.	14	15	16	17	18	19	20	21	22	23	24	25	26	27
	SATELLITE EO METHOD DESIGN															
1a	**Semi-automated pre-processing of Sentinel-2 data for LU/LC Classification**															
	Testing of outputs															
	Documentation															
	Implementation and interactive development															
1b	**Semi-automated pre-processing of Sentinel-1 data for LU/LC Classification**															
	Testing of outputs															
	Documentation															
	Implementation and interactive development															
2	**Annual Forest Biomass Change and Degradation Mapping Using the ALOS PALSAR Mosaic**															
	Calibration and testing of outputs															
	Documentation															
	Implementation and interactive development															
	Validation against available field data															
3	**Continuous Forest Change Monitoring with Sentinel-2 data**															
	Research and development															
	Calibration and testing of outputs															
	Documentation															
	Implementation and interactive development															
	Validation against available field data															
4	**Identifying Causes of Forest Change**															
	Research and development															
	Calibration and testing of outputs															
	Documentation															
	Implementation and interactive development															
	Validation against available field data															

References

Baccini, A.G.S.J., Goetz, S.J., Walker, W.S., Laporte, N.T., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P.S.A., Dubayah, R., Friedl, M.A. and Samanta, S., 2012. Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature Climate Change*, *2*(3), p.182.

Bouvet, A., Mermoz, S., Le Toan, T., Villard, L., Mathieu, R., Naidoo, L. and Asner, G.P., 2018. An above-ground biomass map of African savannahs and woodlands at 25m resolution derived from ALOS PALSAR. Remote *Sensing of Environment*, *206*, pp.156-173.

De Grandi E, Mitchard ETA, Woodhouse IH, De Grandi GD. 2015. Spatial Wavelet Statistics of SAR Backscatter for Characterizing Degraded Forest: A Case Study from Cameroon. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8: 3572-84.

Dong X, Quegan S, Yumiko U, Hu C, Zeng T. 2015. Feasibility Study of C- and L-band SAR Time Series Data in Tracking Indonesian Plantation and Natural Forest Cover Changes. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8: 3692-99

GFOI. 2014. Integrating remote-sensing and ground-based observations for estimation of emissions and removals of greenhouse gases in forests: Methods and Guidance from the Global Forest Observations Initiative, Group on Earth Observations, Geneva

GMV 2016. Generation of a Sentinel-2A Mosaic: Suitable for LULC Classification Purposes.

Hamunyela, E., Verbesselt, J. and Herold, M., 2016. Using spatial context to improve early detection of deforestation from Landsat time series. Remote Sensing of Environment, 172, pp.126-138.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R. and Kommareddy, A., 2013. High-resolution global maps of 21st-century forest cover change. Science, 342(6160), pp.850-853.

Le Toan, T., Quegan, S., Davidson, M.W.J., Balzter, H., Paillou, P., Papathanassiou, K., Plummer, S., Rocca, F., Saatchi, S., Shugart, H., Ulander, L., 2011. The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. Remote Sens. Environ. 115, 2850–2860. doi:10.1016/j.rse.2011.03.020

Le Toan, T., Beaudoin, A., Riom, J., Guyon, D., 1992. Relating forest biomass to SAR data. IEEE Trans. Geosci. Remote Sens. 30, 403–411.

McNicol, I.M., Ryan, C.M., Mitchard, E.T.A., 2018. Carbon losses from deforestation and widespread degradation offset by extensive growth in African woodlands. Nat. Commun. in press.

Mitchard, E.T., Saatchi, S.S., Woodhouse, I.H., Nangendo, G., Ribeiro, N.S., Williams, M., Ryan, C.M., Lewis, S.L., Feldpausch, T.R. and Meir, P., 2009. Using satellite radar backscatter to predict above-ground woody biomass: A consistent relationship across four different African landscapes. Geophysical Research Letters, 36(23).

Motohka, T., Shimada, M., Uryu, Y., Setiabudi, B., 2014. Using time series PALSAR gamma nought mosaics for automatic detection of tropical deforestation: A test study in Riau, Indonesia. Remote Sens. Environ. 155, 79–88. doi:10.1016/j.rse.2014.04.012.

Reiche, J., Hamunyela, E., Verbesselt, J., Hoekman, D. and Herold, M., 2018. Improving near-real time deforestation monitoring in tropical dry forests by combining dense Sentinel-1 time series with Landsat and ALOS-2 PALSAR-2. Remote Sensing of Environment, 204, pp.147-161.

Ryan, C.M., Berry, N.J. and Joshi, N., 2014. Quantifying the causes of deforestation and degradation and creating transparent REDD+ baselines: A method and case study from central Mozambique. *Applied Geography*, *53*, pp.45-54.

Ryan, C.M., Hill, T., Woollen, E., Ghee, C., Mitchard, E., Cassells, G., Grace, J., Woodhouse, I.H., Williams, M., 2012. Quantifying small-scale deforestation and forest degradation in African woodlands using radar imagery. Glob. Chang. Biol. 18, 243–257. doi:10.1111/j.1365-2486.2011.02551.x

Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T., Salas, W., Zutta, B.R., Buermann, W., Lewis, S.L., Hagen, S. and Petrova, S., 2011. Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, *108*(24), pp.9899-9904.

Shimada, M., Ohtaki, T., 2010. Generating large-scale high-quality SAR mosaic datasets: Application to PALSAR data for global monitoring. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 3, 637–656. doi:10.1109/JSTARS.2010.2077619.